Integration of the Rat Recombination and EST Maps in the Rat Genomic Sequence and Comparative Mapping Analysis With the Mouse Genome

Steven P. Wilder,[1,4] Marie-Thérèse Bihoreau,[1,4] Karène Argoud,[1]
Takeshi K. Watanabe,[2] Mark Lathrop,[3] and Dominique Gauguier[1,5]

[1]The Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, Oxford OX3 7BN, UK; [2]Otsuka GEN Research Institute, Tokushima 771-0192, Japan; [3]Centre National de Génotypage, 91006 Evry, France

Inbred strains of the laboratory rat are widely used for identifying genetic regions involved in the control of complex quantitative phenotypes of biomedical importance. The draft genomic sequence of the rat now provides essential information for annotating rat quantitative trait locus (QTL) maps. Following the survey of unique rat microsatellite (11,585 including 1648 new markers) and EST (10,067) markers currently available, we have incorporated a selection of 7952 rat EST sequences in an improved version of the integrated linkage-radiation hybrid map of the rat containing 2058 microsatellite markers which provided over 10,000 potential anchor points between rat QTL and the genomic sequence of the rat. A total of 996 genetic positions were resolved (avg. spacing 1.77 cM) in a single large intercross and anchored in the rat genomic sequence (avg. spacing 1.62 Mb). Comparative genome maps between rat and mouse were constructed by successful computational alignment of 6108 mapped rat ESTs in the mouse genome. The integration of rat linkage maps in the draft genomic sequence of the rat and that of other species represents an essential step for translating rat QTL intervals into human chromosomal targets.

[Supplemental material is available online at www.genome.org.]

The production, assembly, and annotation of the complete sequence of the human genome and more recently genomes of model organisms that are widely used in various fields of biomedical research provide a massive source of information for disease gene discovery. The immediate perspectives of these projects in our understanding of biological processes involved in health and diseases rely on the interpretation of raw sequences in terms of functional annotation and variability among individuals, inbred models, and species (Clamp et al. 2003). As genome sequencing projects initiated in human, mouse, and rat develop (International Human Genome Sequencing Consortium 2001; Waterston et al. 2002; http://www.hgsc.bcm.tmc.edu/projects/rat/), comparative genomics is expected to maximize outcomes of genetic investigations in rodent models towards disease-gene identification in human (Frazer et al. 2003).

Mouse and rat geneticists now possess very similar genetic and genomic tools and resources independently derived for the two species that can participate in disease-susceptibility gene discovery projects and in the functional annotation of the human genome. Essentially, owing to recent progress in the development of gene inactivation methodology in the rat (Zan et al. 2003), the main drawback of rat models is the relatively low number of disease-susceptibility loci mapped in various strains and experimental crosses (http://www.ratmap.gen.gu.se/; http://rgd.mcw.edu/qtls/). On the other hand, comprehensive and accurate screenings of complex phenotypes underlying some of the most frequent and prevalent human genetic disorders, including for example type 2 diabetes and hypertension, have been suc-

cessfully used in rat genetic studies (Gauguier et al. 1996; Stoll et al. 2001), but remain technically challenging in mouse experimental crosses.

In recent years, comparative genome maps between human, mouse, and rat have been rapidly developed in order to facilitate the integration of disease-susceptibility loci identified in the two rodent species and to ultimately provide genetic studies in human with chromosomal targets that can be tested for evidence linkage and association with a disease trait in human populations (Julier et al. 1997; Stoll et al. 2000). The main difficulty in this strategy lies in the fact that rodent genetic loci involved in the control of complex traits are defined by broad intervals in recombination maps cosegregating with variations of a quantitative phenotype in experimental crosses (intercross and backcross) or hybrid populations (recombinant inbred and recombinant congenic panels, heterogeneous stocks), whereas comparative maps are still mainly based on the localization of orthologous genes in physical maps. Although comparative genome analysis will undoubtedly progress with cSNP-based maps and large-scale computational alignments of human, mouse, and rat genomic sequences, the accurate localization of rat disease-susceptibility loci in comparative genome maps still requires full integration of rat genetic and radiation hybrid (RH) maps with rat genome assemblies.

The aim of the present study was to initiate the integration of rat quantitative trait locus (QTL) maps in the rat genomic sequence and in comparative genome maps. We incorporated a selection of 7952 unique rat ESTs in an improved version of the integrated linkage-RH map of the rat containing 2058 microsatellite markers mapped in a single large cross, and used the resulting data for in silico mapping (BLAT and/or electronic-PCR) of both microsatellite markers and ESTs in the emerging rat genomic sequence. We subsequently carried out computational analyses in order to define regions of synteny conservation be-

14:758–765 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

tween the rat and mouse genomes and to fine-map evolutionary breakpoints between the two species in the rat genetic map. Cross-referencing the rat linkage and EST maps to both rat and mouse genome assemblies provides a powerful platform for integrating rat disease-susceptibility loci with functional information generated in mouse models and ultimately applying these resources to human genetics.

RESULTS

Construction of the Genetic Map of the Rat

Our initial objective was to improve marker density in our rat linkage map in order to increase efficiency of electronic polymerase chain reaction (e-PCR) mapping to the rat genomic sequence assembly (http://www.hgsc.bcm.tmc.edu/projects/rat/, release June 2003). The improved version of the autosome recombination map generated in the (GKxBN) F2 cross (Gauguier et al. 1996) now comprises a total of 2058 microsatellites of various collections (mainly DnWox, DnGot, DnRat, DnMco, DnMit, DnMgh, DnUlb sets), including 503 new DnGot microsatellites among the 1607 that could not be assigned to a rat chromosome (RNO) using the T55 RH rat panel (Table 1A; Watanabe et al. 1999). Markers mapped in this cross represent nearly 18% of all 11,585 known microsatellites publicly available (Table 1B), and ~30% of potentially polymorphic markers between GK and BN strains, which exhibit a polymorphism rate of ~60% (Bihoreau et al. 2001). The number of resolved positions (996) increases by 22.5% compared to our previous release of the rat linkage map in the (GKxBN) F2 cross (Bihoreau et al. 2001), without creating a significant elongation of the total linkage map (4%; Table 1A). The average spacing between positions is 1.77 cM. Linkage maps, along with marker information, are available through our rat marker data repository (http://www.well.ox.ac.uk/rat_mapping_resources/).

Anchoring All Known Microsatellites and the Rat Recombination Map to the Rat Genomic Sequence

Primer pairs for all 11,585 known rat microsatellite markers were used to find the most likely localization of these markers in the rat genomic sequence through computational analyses (Schuler 1997). e-PCR, which was solely based on the knowledge of PCR primer sequences and PCR product lengths, allowed the direct assignment of a total of 8097 known microsatellites (69.9%) in the rat genomic sequence (Table 1B). Inconsistent map positions were found for 115 of the markers (1.4%) localized in the rat genomic sequence. Anchor points in the rat genome for all microsatellites that underwent successful e-PCR mapping are publicly available (http://www.ensembl.org/Rattus_norvegicus/; http://www.well.ox.ac.uk/rat_mapping_resources).

A total of 1584 of the 2058 markers mapped in the (GKxBN) F2 cross (77.0%) are now also assigned to the rat genomic sequence, thus allowing direct comparisons between QTL maps and genomic sequence (Table 1A). As shown in Figure 1, marker order in the genetic map and in the genomic sequence is generally well conserved. Inconsistent chromosomal localization between linkage mapping in the GKxBN cross and e-PCR was found for only four (0.3%) markers (D1Wox45, D1Got283, D5Rat18, and D9Got205). Both D1Wox45 and D1Got283 (assigned to RNO17 by e-PCR) were previously localized to RNO1 by genetic and RH mapping (Gauguier et al. 1996; Watanabe et al. 1999). They map less than 1.5 cM apart, and these discrepancies most

probably represent a small error in the draft sequence. Marker D5Rat18 was assigned to RNO3 in our genetic map, although previously localized to RNO5 and remapped to RNO5 by e-PCR, which further supports its localization in RNO5. Marker D9Got205 (assigned to RNO5 by e-PCR) is a novel microsatellite, and thus no further conclusions can be drawn. Details of the position in the genomic sequence for all markers in our recombination map are available through our public database (http://www.well.ox.ac.uk/rat_mapping_resources).

Predicted Localization of Rat EST Sequences in the Linkage Map of the Rat and Integration in the Rat Genomic Sequence

The purpose of remapping rat EST sequences (Scheetz et al. 2001) in our framework RH map was to use our integrated RH-linkage map of the rat (Bihoreau et al. 2001) as a tool for anchoring rat QTL maps in comparative genome maps. Using publicly available RH scores for a total of 10,067 unique EST sequences (http://corba.ebi.ac.uk/Rhdb) defined by nonredundant primer pairs, we were able to assign 7952 of these sequences (79%) in our RH-framework map of the rat autosomes (Table 2) and subsequently into genetic intervals or positions of our integrated RH-linkage map of the rat (Bihoreau et al. 2001). Approximately 20% of the ESTs were localized at a single genetic position, whereas the remaining were assigned to most likely genetic intervals of the recombination map (Table 2). Only 303 (3.8%) inconsistencies in chromosomal localization, mainly in chromosome extremities, were observed with the published localization (Scheetz et al. 2001). All EST mapping data, including problematic data, are available in our public database (http://www.well.ox.ac.uk/rat_mapping_resources).

All 7952 rat ESTs that have been used to generate RH scores in the rat T55 panel were searched for homologous sequences in

Table 1A. Outlined Description of the Genetic Map in the (GK × BN) F2 Cross and Integration of Linkage and In Silico Maps of the Rat Autosomes

RNO	Markers[a]	Resolved positions	Average (cM)[a]	Average (Mb)[a]	In silico mapping[b]	Average (Mb)[b]
1	333	103	1.52	2.63	263	1.02
2	192	90	1.26	2.90	150	1.73
3	110	64	1.97	2.71	91	1.90
4	159	69	1.67	2.76	115	1.64
5	172	69	1.45	2.55	132	1.32
6	102	48	1.47	3.14	81	1.85
7	104	58	1.66	2.51	77	1.88
8	108	56	1.64	2.35	86	1.52
9	75	41	2.11	2.84	52	2.23
10	128	92	1.57	1.22	98	1.14
11	40	24	2.94	3.82	34	2.66
12	50	26	1.95	1.87	34	1.41
13	78	33	2.02	3.48	55	2.06
14	72	34	2.15	3.40	53	2.16
15	53	34	2.63	3.33	44	2.55
16	45	24	2.78	3.92	36	2.58
17	92	44	1.54	2.26	78	1.26
18	59	31	1.81	2.91	44	2.03
19	41	26	1.98	2.37	30	2.04
20	45	30	2.41	1.91	31	1.84
Total	2058	996	1.77	2.57	1584	1.62

[a]Number of microsatellite markers mapped in the (GK × BN) F2 cross and average spacing between resolved genetic positions. Average physical distances were calculated with the chromosome length determined in the June 2003 rat genome assembly.
[b]Markers localized in the genetic map (GK × BN) F2 cross that could be mapped by ePCR and average physical distances calculated as described above.

Table 1B. Description of In Silico Assignment of Microsatellite Markers to the Rat Genome

RNO	Markers[a]	e-PCR mapping[a]	Average spacing (kb)[b]	Inconsistencies (%)[c]
1	1149	834	293.7	26 (3.1%)
2	948	632	371.0	6 (0.9%)
3	675	497	303.7	10 (2.0%)
4	692	470	352.9	2 (0.4%)
5	694	475	323.6	3 (0.6%)
6	540	403	348.2	14 (3.5%)
7	573	414	316.6	7 (1.7%)
8	580	425	277.0	12 (2.8%)
9	453	310	335.3	5 (1.6%)
10	686	480	215.4	5 (1.0%)
11	284	196	397.3	3 (1.5%)
12	307	205	215.0	4 (2.0%)
13	476	316	311.9	4 (1.3%)
14	355	258	377.8	3 (1.2%)
15	381	266	375.9	4 (1.5%)
16	315	223	372.8	2 (0.9%)
17	416	311	281.2	2 (0.6%)
18	325	234	342.5	2 (0.9%)
19	253	172	297.6	0 (0.0%)
20	194	123	389.4	1 (0.8%)
Unassigned	1289	853		
Total	11,585	8097	270.4	115 (1.4%)

[a]The number of markers used for e-PCR mapping to the rat genome sequence includes all available microsatellites, including 1289 markers previously defined by a D-Number nomenclature (e.g., D0Got, D0Mco) that could not be assigned to a rat chromosome by genetic or RH mapping.
[b]The average distance between markers in the genomic sequence was calculated with the chromosome length determined in the June 2003 rat genome assembly, and includes above-described markers that are only mapped by e-PCR.
[c]Details of markers that are assigned to different chromosomes by linkage or RH mapping and e-PCR are available in our data repository (http://www.well.ox.ac.uk/rat_mapping_resources).

the rat genomic sequence (June 2003) using automated BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat). Of the 7952 ESTs integrated in the rat recombination map, a total of 7224 (90.8%) could also be localized in the rat genomic sequence, with only 333 discordant mappings between RH and BLAT (Table 2). Many of the ESTs were mapped to the genomic sequence independently by BLAT and e-PCR. When both methods produced a single significant map position, there was very high agreement between them [five out of 4943 ESTs (0.1%) were mapped to different chromosomes by these methods], and of all those mapped to the same chromosome; the maximum distance between positions was 69 kb (data not shown), verifying the use of either method of mapping using the rat genomic sequence.

EST Sequence Homology Searches in the Mouse Genome and Comparative Mapping Analysis

We subsequently carried out a BLAT search using the complete set of 7952 rat EST sequences which were mapped against our RH map (Watanabe et al. 1999), and identified 6108 homologous sequences in the mouse genome (October 2003; Table 2). Among these, 5688 (93.1%) were also uniquely mapped in the rat genome. These mapping data were used to identify evidence of conservation between the mouse and rat genomes and localize both intrachromosomal rearrangements in the rat and breakpoints of synteny conservation between the two species. Comparative genome analysis derived from EST mapping generally confirms previously identified conserved chromosomal regions

in mouse and rat (Fig. 1; Watanabe et al. 1999; Bihoreau et al. 2001; Helou et al. 2001; Kwitek et al. 2001). Only a few previously defined rat–mouse chromosomal homologies occurring in either centromeric (RNO5, 7, 12, 13, 18, and 19) or telomeric (RNO10, 15, and 18) regions (Helou et al. 2001; Kwitek et al. 2001) were not confirmed in our study. Graphs for all autosomes are available at http://www.well.ox.ac.uk/rat_mapping_resources/.

The definition of anchor points in the two genomic sequences allowed the characterization of 39 synteny groups and 45 interchromosomal rearrangements (Table 2). To illustrate examples of full synteny conservation between rat and mouse and fragmented synteny conservation interrupted by evolutionary breakpoints, graphical representations of the comparative mapping are shown for RNO 1, 8, and 20 (Fig. 1). Idiograms of these same chromosomes are depicted, indicating the correspondence between the chromosomal genetic and physical maps, and the syntenic conservation between the rat and mouse genomes (Fig. 2). Intrachromosomal rearrangements can be preliminarily identified by eye with cat's-cradle patterns in the comparative mapping figures, as observed for RNO20 (Fig. 2). An inversion in the conserved region is represented by a small number of points forming a line, in the direction opposite to the primary direction of conservation (Fig. 1). Insertions/deletions are represented as either horizontal or vertical discontinuities in the line of syntenic conservation. For example, an insertion in the rat genomic sequence relative to that of the mouse would be depicted as a vertical gap (Fig. 1).

DISCUSSION

The integration of the rat genetic map in the emerging rat and mouse genomic sequences reported here provides an initial step towards a comprehensive functional annotation of rat QTL for human complex traits. Increased microsatellite marker density in our rat linkage map followed by e-PCR allowed the definition of 1584 anchor points in the rat genome sequence, spaced by an average of 1.62 Mb. Such anchoring enabled subsequent rat–mouse comparative genome analysis using EST and gene sequences localized in our integrated RH-linkage map of the rat.

The importance of inbred strains of the laboratory rat for genetic investigations of human complex traits was addressed in detail by Jacob and Kwitek (2001). Genetic studies in the rat are essentially driven by the wealth of physiological and pathophysiological information that can be quantified in large experimental crosses (Gauguier et al. 1996; Stoll et al. 2001). Elucidation of the genetic basis of complex phenotypes in the rat has now successfully evolved from QTL mapping in experimental crosses to congenic-based strategies that allowed the identification of susceptibility genes for insulin resistance, diabetes mellitus, and inflammation (Aitman et al. 1999; Fakhrai-Rad et al. 2000; Marion et al. 2002; Olofsson et al. 2003). Owing to the comprehensive and accurate phenotypic characterization that can be achieved in rat congenic lines (Rogner and Avner 2003), improved functional annotation of the chromosomal regions targeted in congenics will provide essential insights into the biological role of genetic alterations responsible for the QTL effects. Integrating genetic maps and genomic sequences provides direct links between QTL and congenic intervals and functional information derived from genome annotation. In return, predicted marker position in the genetic map and QTL intervals can further improve the characterization of chromosomal segments introgressed in congenic lines.

Our improved genetic map of the rat, constructed in a single large intercross, now comprises over 2000 microsatellite markers, representing nearly 20% of all previously known markers and 30% of potentially polymorphic markers in a GKxBN strain com-
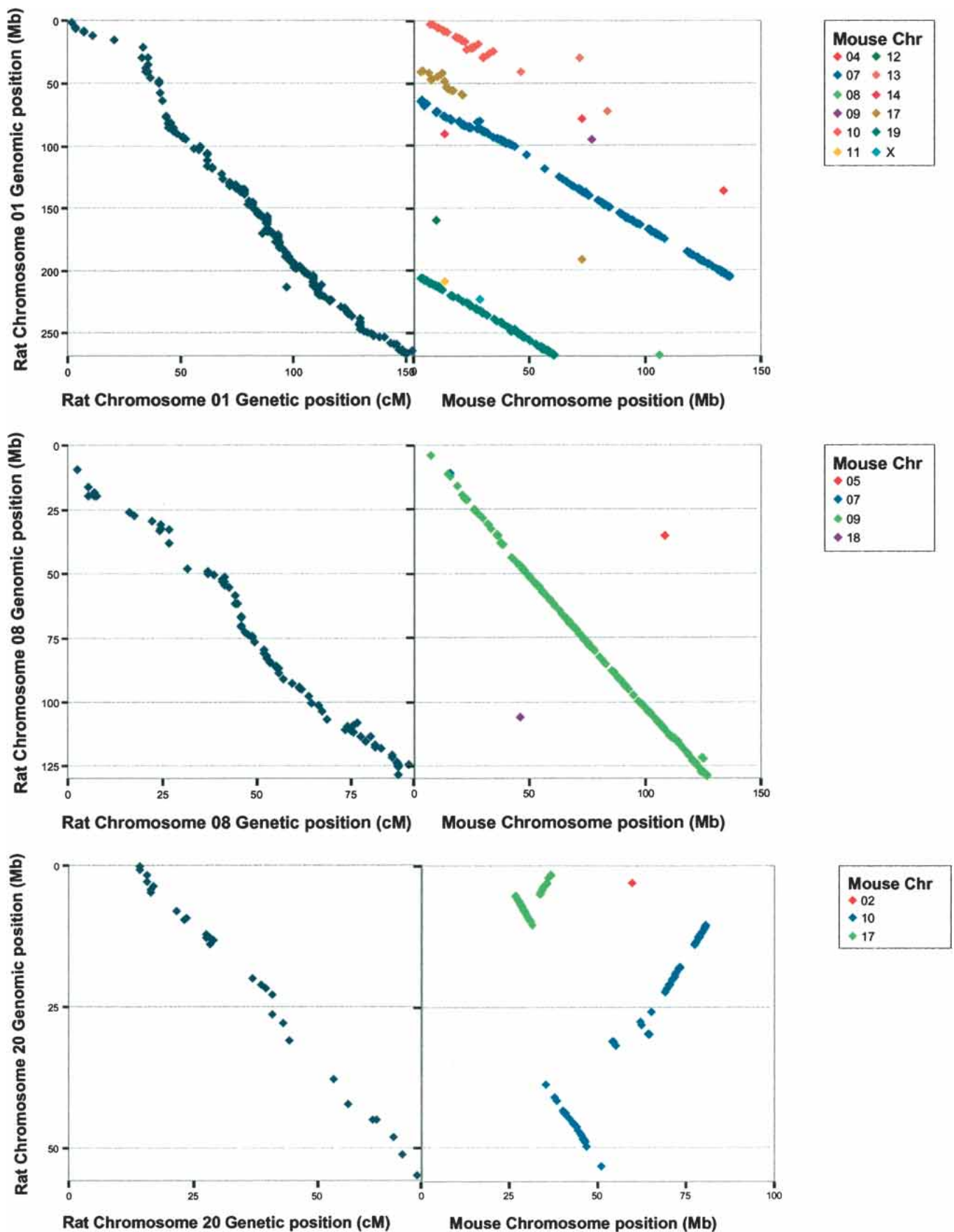
Figure 1 (Legend on next page)

Table 2. Integration of the ESTs in the RH and Linkage Maps of the Rat Autosomes and Synteny Conservation With the Mouse Genome

		RH mapping			Resolved genetic positions[d]	Rat genome	Sequence homology (BLAT)[e]			Synteny groups
RNO	Total[a]	Mapped[b]	Discordant mapping	Resolved positions[c]			Discordant RH-BLAT	Mouse genome	Both genomes	
1	1320	1095	58	271	192	1014	63	793	751	5
2	775	537	37	201	178	491	21	415	382	3
3	598	414	10	174	135	379	13	331	313	1
4	757	443	43	98	53	411	12	356	339	2
5	620	501	40	145	106	466	27	395	374	1
6	429	406	26	103	83	366	40	303	277	3
7	604	550	4	145	107	485	22	423	387	3
8	606	560	17	172	112	496	16	448	412	1
9	341	287	5	75	44	259	11	212	198	2
10	723	619	2	152	118	565	17	501	466	3
11	308	256	35	77	34	237	6	195	183	1
12	281	270	5	81	40	251	16	186	175	1
13	324	310	5	78	32	273	8	257	233	1
14	271	254	2	65	35	232	11	179	169	2
15	381	239	12	59	35	213	14	186	168	1
16	321	286	1	65	24	260	7	229	213	2
17	604	260	0	63	46	235	4	198	185	3
18	298	247	0	55	47	217	12	188	174	1
19	230	163	0	38	37	152	6	126	120	1
20	276	255	0	54	48	222	7	187	169	2
Total	10,067	7952	303	2171 (27%)	1506 (19%)	7224	333	6108	5688	39

[a]Number of ESTs whose RH scores were extracted from the RHdb web site (http://corba.ebi.ac.uk/Rhdb).
[b]Number of ESTs successfully mapped to our RH framework maps (Watanabe et al. 1999).
[c]ESTs which map to resolved positions defining framework RH positions (Watanabe et al. 1999).
[d]ESTs which map to the resolved positions defining both framework positions of the RH map (Watanabe et al. 1999) and genetic loci in the linkage map derived from the BN × GK intercross (Gauguier et al. 1996).
[e]EST markers mapped reliably to the rat and mouse genomic sequences using BLAT (Kent 2002). Details of ESTs that are assigned to different chromosomes by RH mapping and BLAT are available in our data repository (http://www.well.ox.ac.uk/rat_mapping_resources).

bination. Increased marker density in the linkage map led to a significant improvement in average spacing between markers (1.77 cM) compared to our previous map (2.26 cM; Bihoreau et al. 2001). Out of the 11,585 unique primer pairs that we used to annotate the rat genomic sequence for known microsatellite markers, we identified unambiguous genome localization for 8097 markers, including 1584 markers already assigned to a chromosome in our recombination map. The apparently high failure rate (30%) is primarily due to mismatches in the primer sequences, caused by errors in the genomic sequence or polymorphisms. We required that primer sequences had at most one base pair mismatch in the genomic sequence, and that the PCR product length calculated by e-PCR on the genome was within 50 base pairs of the expected length. Conservation in marker order between linkage and genomic sequence was generally observed. In this respect, the draft sequence of the rat genome provides an opportunity for clarifying marker order in chromosomal regions that lack genetic mapping resolution. By comparing the genetic distance from our linkage map with the physical distances generated by the e-PCR program, there is again evidence of the non-linear relationship between them (Fig. 1). We could estimate over the genome the correspondence between physical and genetic distance as 1.50 Mb/cM. This is consistent with our estimates in a previous study, which reported the evaluated correspondence to be 16.3 cR_{3000}/cM and 1.73 Mb/cM (Bihoreau et al. 2001).

Evidence of synteny conservation and disruption among the rat, mouse, and human genomes has already been investigated in detail (Gauguier et al. 1999; Watanabe et al. 1999; Kaisaki et al. 2000; Kwitek et al. 2001). Previous comparative genome analyses based on rat gene mapping showed that, although synteny conservation between rat and mouse genomes is extensive, homologs of a limited proportion of annotated rat genes (<40%) can be found in mouse genome databases, whereas ~70% are also annotated in human (Bihoreau et al. 2001). These estimates suggested that gene mapping data derived in the rat can substantially enrich existing comparative gene maps between human and mouse. The availability of the rat genomic sequence assembly provides crucial information for integrating functional annotations of rat and mouse QTL, translating disease loci from rodent models to human, and investigating structures involved in genome evolution. Comparative genome analysis will have a strong impact on the functional annotation of rat QTL and congenic intervals in two important ways. First, access to genomic sequences of various species in the region of a rat disease-susceptibility locus will provide information on known genes, pseudogenes, and noncoding sequences (Clamp et al. 2003). Sec-

Figure 1 Examples of correlation between genetic and genomic positions and comparative analysis between rat (RNO) and mouse (MMU) genomes for RNO 1, 8, and 20. Each point corresponds to a rat microsatellite marker (e-PCR) or a rat EST sequence (BLAT). The genomic localization of the microsatellite markers assigned to RNO 1, 8, and 20 linkage maps derived from the GKxBN intercross was determined by the e-PCR method (*left*-side panels). Lengths are in cM (Kosambi) in the linkage map and Mb in the genomic sequence. Chromosomal synteny conservation was defined by BLAT query of rat EST sequences (*right*-side panels). The orientation of the rat chromosomes is as established by Szpirer et al. (1998). Graphs for all autosomes are available in the Supplemental material and at http://www.well.ox.ac.uk/rat_mapping_resources.
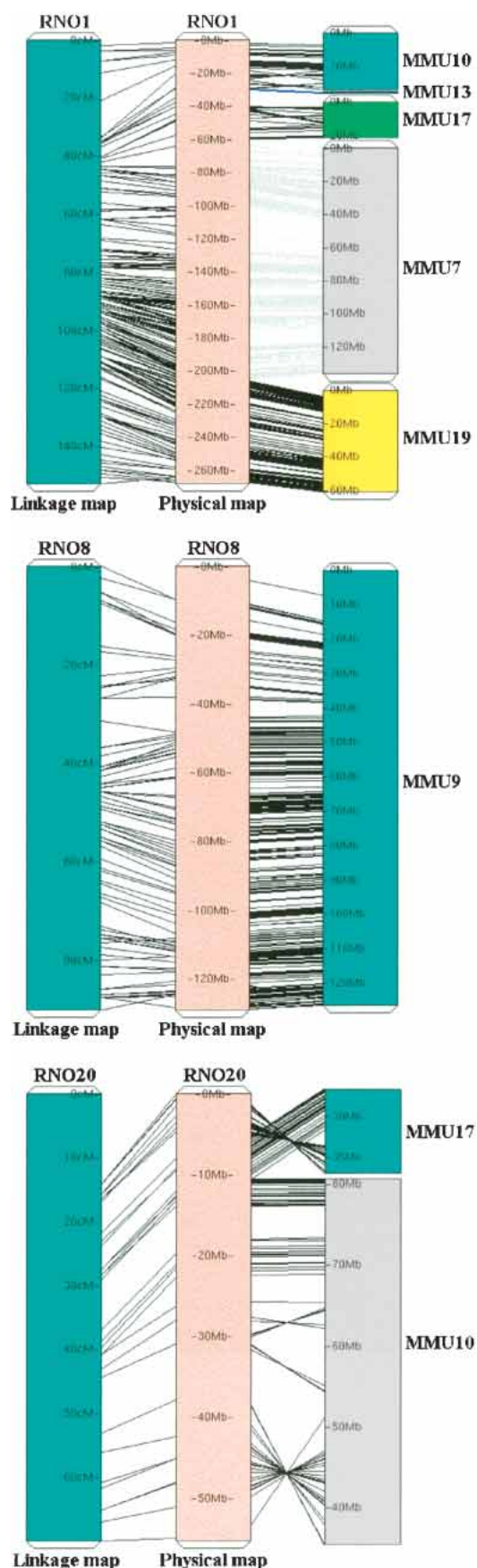
ond, fine mapping of evolutionary breakpoints between mammalian genomes will identify human chromosomal regions that are directly homologous to a rat disease-susceptibility locus.

Mapping rat EST sequences in our rat RH framework map and inferring their position in our rat recombination map was a preliminary step towards the integration of potential gene sequences in the rat and mouse genomic sequences. The present study was limited to sequence homology searches between rat and mouse ESTs, as it is anticipated that the construction of mouse–human comparative maps, which is well under way (Waterston et al. 2002), will provide essential links for rat–human genome relationships. With our approach, we could define 39 groups of synteny conservation between the rat and mouse genomes, and 45 contiguous conserved regions as compared to 51, 59, 49, or 39 resulting from the RH mapping of ESTs (Kwitek et al. 2001), the RH mapping of gene markers (Watanabe et al. 1999), the mouse-on-rat zoo-FISH analysis of orthologous rat–mouse gene pairs (Gomez-Fabre et al. 2002), and the linkage mapping of microsatellites associated to genes (Bihoreau et al. 2001), respectively. The combination of these data gives a consensus of 33 segments of synteny conservation between rat and mouse.

Although the comparative genome analysis reported here largely supports known rat–mouse homology relationships, some differences became apparent, and we sought to ascertain which position was likely to be correct by checking the data against other publicly available maps, for example, cytogenetic map (Szpirer et al. 1998), Rat Genome Database (RGD) genetic map, and RH map (Steen et al. 1999; Kwitek et al. 2001). The majority of previously reported rat–mouse segments of synteny conservation that we did not confirm in this study occur in either centromeric or telomeric regions of rat chromosomes. This may be due to a higher rate of substitutions/insertions at the extremes of chromosomes, and the paucity of markers in these regions, or inaccuracies in these parts of the published genomic sequence.

As each EST in the study has an interval in the genetic map assigned to it through the RH framework map, one can aim to refine the rat-to-mouse syntenic breakpoints mapping on both the genetic and physical map. In some cases we can locate these breakpoints to within 0.1 Mb or 2 cM. For example, on RNO6, AI071025 is mapped at position 24.03 Mb and comparatively mapped to MMU17, whereas AI555529 is mapped at 24.07 Mb on RNO6 but aligns with MMU5, and so the breakpoint between conserved syntenic segments can be narrowed down to this 40 kb region (http://www.ensembl.org/Rattus_norvegicus). As both of these ESTs are in the same interval on the integrated RH-linkage framework map, we can position the breakpoint between 10.6 and 11.2 cM on RNO6.

Overall, results reported here provide important information for improving the functional annotation of genomic sequences underlying rat and murine QTL and identifying candidate genes for human complex traits. It is anticipated that the full annotation of rat and mouse genomes will be synergistic for the identification of disease genes in human. Comparative genome maps, which were until now based on the chromosomal

Figure 2 Idiograms for rat RNO 1, 8, and 20. The *left*-side block indicates the chromosomal linkage map derived from the GK × BN intercross; the *central* block indicates the chromosomal physical map. The lines between these two blocks represent the genetic position of a microsatellite and its physical position as determined by e-PCR. The *right*-side blocks indicate the suggested respective homologous segments of the mouse genome, with the white trapeziums indicating the ends of the mouse chromosomes. The lines on the *right* side of the figure indicate the comparative mapping of ESTs between the rat and mouse genomes as determined by BLAT searches. Cat's-cradle patterns identify intrachromosomal rearrangements between the two genomes.

mapping of known genes and putative mRNA, will ultimately progress with sequence alignments allowing genome-wide analysis of noncoding sequences (Cooper et al. 2003; Frazer et al. 2003) and definition of homology and diversity genomic patterns between species.

METHODS

Microsatellite Markers

A total of 11,585 microsatellite markers of various origins, for which primer pairs and PCR product length were known, were used for genetic mapping and/or in silico localization in the rat genomic sequence. These included large series of 1648 novel microsatellites (DnGot markers) generated by Otsuka laboratories (Watanabe et al. 2000). Primer pairs for these markers generally amplified PCR products of similar size on rat and hamster DNA, which did not allow chromosomal mapping using the T55 rat/hamster radiation hybrid panel (Watanabe et al. 1999). PCR primers for new microsatellite markers were synthesized commercially by Sigma-Genosys Biotechnologies. Information on the survey of all microsatellite markers used in this study is publicly available (http://www.well.ox.ac.uk/rat_mapping_resources/). The database has been carefully curated to identify potential duplicates.

Genetic Mapping of New Microsatellite Loci

A single large intercross derived from the Goto-Kakizaki (GK) and Brown Norway (BN) rats (n=139) was used to integrate new DnGot microsatellite markers showing allele variations between the two parental strains into existing linkage maps (Gauguier et al. 1996, 1999; Kaisaki et al. 2000; Bihoreau et al. 2001). Genotyping was performed as described (Bihoreau et al. 2001).

Construction of the Linkage Maps

Linkage maps were constructed using the JoinMap version 2.0 suite of programs (Stam 1995) as described (Bihoreau et al. 1997). Briefly, single factor segregation ratios were calculated for each marker in a linkage group and, following verification of double recombination events, genetic maps were created. The JoinMap module for genotype checking calculates the probability of obtaining the present genotype for all loci and for all individuals, conditional on both the genotypes at the two flanking loci and map distances.

In Silico Assignment of Rat Microsatellite Loci in the Rat Genomic Sequence

Microsatellites were mapped into the rat genomic sequence by electronic PCR (e-PCR; Schuler 1997), allowing one mismatch in the primer sequence and an error of ±50 in the expected PCR product length. Multiple hits were accepted only if all hits were within the same 50-kb region, in which case the mean position was used. The e-PCR program works by aligning a microsatellite to an interval in the genomic sequence with the two primers in the correct configuration (i.e., the sequence of one of the primers is reversed), with a gap between them of approximately the specified PCR product length. Many of the microsatellites have been mapped previously in our and other linkage maps and radiation hybrid (RH) maps.

Integration of Rat EST Loci in the Integrated RH-Linkage Map of the Rat

Publicly available RH scores generated on the T55 rat RH panel for rat EST sequences were retrieved from public databases (http://corba.ebi.ac.uk/Rhdb) and used to map the EST loci against our RH framework map (Watanabe et al. 1999) using stringent criteria. Two-point analysis was initially carried out, and the only markers considered for multipoint analysis were those supported both by a two-point lod score of at least 6 with a framework RH marker, and with the three highest two-point

linkages within a contiguous region of the same chromosome. This robust analysis allowed either the integration of EST sequences at a single RH position or the identification of the best interval between RH framework positions.

Comparative Genome Analysis

Rat EST loci localized in our RH framework map were used to search for homologous loci in the mouse genome sequence. The sequence for each rat EST, obtained from a public database (EMBL), was masked for rodent repeats using RepeatMasker (http://repeatmasker.genome.washington.edu/). These masked sequences were then aligned against both the rat and mouse genomic sequences, using BLAT (Kent 2002). BLAT was chosen in preference to BLAST (Altschul et al. 1990), as BLAT runs much faster when comparing the query sequence to an entire mammalian genome, and because the evolutionary distance between rat and mouse is small, distant homologs which may be missed using BLAT are unlikely. Both programs were run on a small subset of the EST sequences producing near-identical primary hits (data not shown). The BLAT results were then automatically processed in order to remove any low-scoring matches or ESTs that were mapped with similar significance to multiple points in different areas of the genome, indicating repeats that were not recognized by RepeatMasker.

The criteria for a hit from the BLAT query to define a potential homolog were based on the score of the highest-scoring alignment compared to the other significant scoring alignments, and their relative position in the genome. If the two map positions were close enough to indicate a local repeat, the primary hit was accepted. If the primary hit scored substantially higher than any subsequent hits, the first mapping position was accepted. All EST sequences whose BLAT mapping position passed these criteria were then considered potential homologs, and were used in the comparative mapping of the rat and mouse genomes. To prevent false positive matches, only segments of synteny conservation on the mouse genome containing at least two distinct ESTs were accepted. Therefore our prediction for the number of syntenic segments is likely to be an underestimate.

Finally, each of the rat ESTs is contained in our RH framework map and all of the framework endpoints are microsatellites, the majority of which were assigned to positions in the rat genomic sequence by e-PCR. Hence by merging these two maps, we could assign an interval on our genetic map to all of the ESTs. Information on all rat ESTs used in this study for RH mapping and/or comparative mapping is available through our public database (http://www.well.ox.ac.uk/rat_mapping_resources/).

REFERENCES

Aitman, T.J., Glazier, A.M., Wallace, C.A., Cooper, L.D., Norsworthy, P.J., Wahid, F.N., Al-Majali, K.M., Trembling, P.M., Mann, C.J., Shoulders C.C., et al. 1999. Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nat. Genet.* **21:** 76–83.
Altschul, S., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990.

Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Bihoreau, M.T., Gauguier, D., Kato, N., Hyne, G., Lindpaintner, K., Rapp, J.P., and Lathrop, G.M. 1997. A linkage map of the rat genome derived from three F2 crosses. *Genome Res.* **7:** 434–440.

Bihoreau, M.T., Sebag-Montefiore, L., Godfrey, R.F., Wallis, R.H., Brown, J.H., Danoy, P.A., Collins, S.C., Rouard, M., Kaisaki, P.J., Lathrop, M., et al. 2001. A high-resolution consensus linkage map of the rat, integrating radiation hybrid and genetic maps. *Genomics* **75:** 57–69.

Clamp, M., Andrews, D., Barker, D., Bevan, P., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., et al. 2003. Ensembl 2002: Accommodating comparative genomics. *Nucleic Acids Res.* **31:** 38–42.

Cooper, G.M., Brudno, M., Green, E.D., Batzoglou, S., and Sidow, A. 2003. Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Res.* **13:** 813–820.

Fakhrai-Rad, H., Nikoshkov, A., Kamel, A., Fernstrom, M., Zierath, J.R., Norgren, S., Luthman, H., and Galli, J. 2000. Insulin-degrading enzyme identified as a candidate diabetes susceptibility gene in GK rats. *Hum. Mol. Genet.* **9:** 2149–2158.

Frazer, K.A., Elnitski, L., Church, D.M., Dubchak, I., and Hardison, R.C. 2003. Cross-species sequence comparisons: A review of methods and available resources. *Genome Res.* **13:** 1–12.

Gauguier, D., Froguel, P., Parent, V., Bernard, C., Bihoreau, M.T., Portha, B., Pénicaud, L., Lathrop, M., and Ktorza, A. 1996. Chromosomal mapping of genetic loci associated with non-insulin dependent diabetes in the GK rat. *Nat. Genet.* **12:** 38–43.

Gauguier, D., Kaisaki, P.J., Rouard, M., Wallis, R.H., Browne, J., Rapp, J.P., and Bihoreau, M.T. 1999. A gene map of the rat derived from linkage analysis and related regions in the mouse and human genomes. *Mamm. Genome* **10:** 675–686.

Gomez-Fabre, P.M., Helou, K., and Stahl, F. 2002. Predictions based on the rat–mouse comparative map provide mapping information on over 6000 new rat genes. *Mamm. Genome.* **13:** 189–193.

Helou, K., Walentinsson, A., Levan, G., and Stahl, F. 2001. Between rat and mouse zoo-FISH reveals 49 chromosomal segments that have been conserved in evolution. *Mamm. Genome.* **12:** 765–771.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Jacob, H.J. and Kwitek, A.E. 2001. Rat genetics: Attaching physiology and pharmacology to the genome. *Nat. Rev. Genet.* **3:** 33–42.

Julier, C., Delepine, M., Keavney, B., Terwilliger, J., Davis, S., Weeks, D.E., Bui, T., Jeunemaitre, X., Velho, G., Froguel, P., et al. 1997. Genetic susceptibility for human familial hypertension in a region of homology with blood pressure linkage on rat chromosome 10. *Human Mol. Genet.* **6:** 2077–2085.

Kaisaki, P.J., Rouard, M., Danoy, P.A.C., Wallis, R.H., Collins, S.C., Rice, M., Levy, E.R., Lathrop, M., Bihoreau, M.T., and Gauguier, D. 2000. Detailed comparative gene map of rat chromosome 1 with mouse and human genomes and physical mapping of an evolutionary chromosomal breakpoint. *Genomics* **64:** 32–43.

Kent, W. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* **12:** 656–664.

Kwitek, A.E., Tonellato, P.J., Chen, D., Gullings-Handley, J., Cheng, Y.S., Twigger, S., Scheetz, T.E., Casavant, T.L., Stoll, M., Nobrega, M.A., et al. 2001. Automated construction of high-density comparative maps between rat, human, and mouse. *Genome Res.* **11:** 1935–1943.

Marion, E., Kaisaki, P.J., Pouillon, V., Gueydan, C., Levy, J., Bodson, A., Krzentowski, G., Daubresse, J.C., Mockel, J., Behrends, J., et al. 2002. The gene INPPL1, encoding the lipid phosphatase SHIP2, is a candidate for type 2 diabetes in rat and man. *Diabetes* **51:** 2012–2017.

Olofsson, P., Holmberg, J., Tordsson, J., Lu, S., Akerstrom, B., and Holmdahl, R. 2003. Positional identification of Ncf1 as a gene that regulates arthritis severity in rats. *Nat. Genet.* **33:** 25–32.

Rogner, U.C. and Avner, P. 2003. Congenic mice: Cutting tools for complex immune disorders. *Nat. Rev. Immunol.* **3:** 243–252.

Scheetz, T.E., Raymond, M.R., Nishimura, D.Y., McClain, A., Roberts, C., Birkett, C., Gardiner, J., Zhang, J., Butters, N., Sun, C., et al. 2001. Generation of a high-density rat EST map. *Genome Res.* **11:** 497–502.

Schuler, G.D. 1997. Sequence mapping by electronic PCR. *Genome Res.* **7:** 541–550.

Stam, P. 1995. Construction of integrated genetic linkage maps by means of a computer package: JoinMap. *Plant J.* **5:** 739–744.

Steen, R.G., Kwitek Black, A.E., Glenn, C., Gullings Handley, J., Van Etten, W., Atkinson, O.S., Appel, D., Twigger, S., Muir, M., Mull, T., et al. 1999. A high-density integrated genetic linkage and radiation hybrid map of the laboratory rat. *Genome Res.* **9:** 1–8.

Stoll, M., Kwitek-Black, A.E., Cowley Jr., A.W., Harris, E.L., Harrap, S.B., Krieger, J.E., Printz, M.P., Provoost, A.P., Sassard, J., and Jacob, H.J. 2000. New target regions for human hypertension via comparative genomics. *Genome Res.* **10:** 473–482.

Stoll, M., Cowley Jr., A.W., Tonellato, P.J., Greene, A.S., Kaldunski, M.L., Roman, R.J., Dumas, P., Schork, N.J., Wang, Z., and Jacob, H.J. 2001. A genomic-systems biology map for cardiovascular function. *Science* **294:** 1723–1726.

Szpirer, C., Szpirer, J., Van Vooren, P., Tissir, F., Simon, J.S., Koike, G., Jacob, H.J., Lander, E.S., Helou, K., Klinga-Levan, K., et al. 1998. Gene-based anchoring of the rat genetic linkage and cytogenetic maps: New regional localizations, orientation of the linkage groups, and insights into mammalian chromosome evolution. *Mamm. Genome* **9:** 721–734.

Watanabe, T.K., Bihoreau, M.T., McCarthy, L.C., Kiguwa, S.L., Hishigaki, H., Tsuji, A., Browne, J., Yamasaki, Y., Mizoguchi-Miyakita, A., Oga, K., et al. 1999. A map of the rat genome containing 5,203 markers: 4,700 microsatellites and 605 genes in a rat, mouse and human comparative map. *Nat. Genet.* **2:** 27–36.

Watanabe, T.K., Ono, T., Okuno, S., Mizoguchi-Miyakita, A., Yamasaki, Y., Kanemoto, N., Hishigaki, H., Oga, K., Takahashi, E., Irie, Y., et al. 2000. Characterization of newly developed SSLP markers for the rat. *Mamm. Genome* **11:** 300–305.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562.

Zan, Y., Haag, J.D., Chen, K.S., Shepel, L.A., Wigington, D., Wang, Y.R., Hu, R., Lopez-Guajardo, C.C., Brose, H.L., Porter, K.I., et al. 2003. Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat. Biotechnol.* **21:** 645–651.

WEB SITE REFERENCES

http://genome.ucsc.edu/cgi-bin/hgBlat; BLAT Search Genome, UCSC Genome Browser.

http://corba.ebi.ac.uk/Rhdb; Radiation hybrid data repository, European Bioinformatics Institute.

http://www.ensembl.org/Rattus_norvegicus; Rat genome annotation, European Bioinformatics Institute.

http://www.hgsc.bcm.tmc.edu/projects/rat; Rat Genome Project, Baylor College of Medicine.

http://rgd.mcw.edu/qtls; Rat genome database, Medical College of Wisconsin, Milwaukee.

http://repeatmasker.genome.washington.edu; RepeatMasker Server, University of Washington.

http://www.ratmap.gen.gu.se; Rat genome database RatMap, Göteborg University.

http://www.well.ox.ac.uk/rat_mapping_resources; Rat genetic data repository, Wellcome Trust Centre for Human Genetics, University of Oxford.